MS-E2122 - Nonlinear Optimization
Lecture XI

Fernando Dias

Department of Mathematics and Systems Analysis

Aalto University
School of Science

November 8, 2023

Outline of this lecture

Barrier functions

Barrier method

Interior point method for LP/ QP

The end... ?

Outline of this lecture

Barrier functions

Barrier method

Interior point method for LP/ QP

The end... ?

Barrier functions

Same idea as in penalty methods: turn constrained optimisation unconstrained and solve them iteratively.

Main difference: barrier functions prevent the search from leaving the feasible region. Consider the primal problem P

$$\begin{aligned}(P): \ \text{min.} \ & f(x) \\ \text{subject to:} \ & g(x) \leq 0 \\ & x \in X.\end{aligned}$$

Barrier functions

Same idea as in penalty methods: turn constrained optimisation unconstrained and solve them iteratively.

Main difference: barrier functions prevent the search from leaving the feasible region. Consider the primal problem P

$$\begin{aligned}(P):\ \text{min.}\ & f(x)\\ \text{subject to:}\ & g(x) \leq 0\\ & x \in X.\end{aligned}$$

We define the barrier problem BP as

$$\begin{aligned}(BP):\ & \inf_{\mu}\ \theta(\mu)\\ \text{subject to:}\ & \mu > 0,\end{aligned}$$

where $\theta(\mu) = \inf_x \{f(x) + \mu B(x) : g(x) < 0, x \in X\}$ and $B(x)$ is a barrier function.

Barrier functions

The barrier function $B : \mathbb{R}^m \to \mathbb{R}$ is such that

$$B(x) = \sum_{i=1}^{m} \phi(g_i(x)), \text{ where } \begin{cases} \phi(y) \geq 0, & \text{if } y < 0; \\ \phi(y) = \infty, & \text{when } y \to 0^-. \end{cases} \quad (1)$$

Barrier functions

The barrier function $B : \mathbb{R}^m \to \mathbb{R}$ is such that

$$B(x) = \sum_{i=1}^{m} \phi(g_i(x)), \text{ where } \begin{cases} \phi(y) \geq 0, & \text{if } y < 0; \\ \phi(y) = \infty, & \text{when } y \to 0^-. \end{cases} \tag{1}$$

Some common alternatives include

- $B(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)}$
- $B(x) = -\sum_{i=1}^{m} \ln(\min\{1, -g_i(x)\})$.

Perhaps the most important is Frisch's log barrier function

$$B(x) = -\sum_{i=1}^{m} \ln(-g_i(x)).$$

Barrier functions

Ideally, $B(x)$ would serve as an indicator function

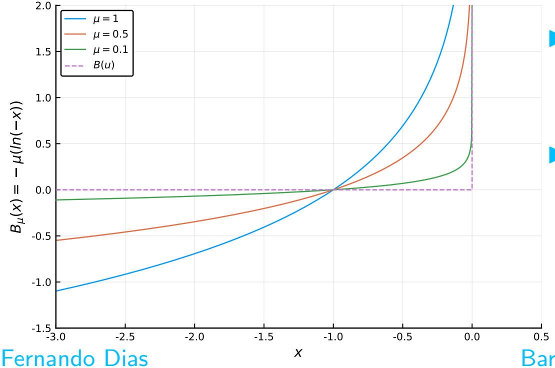
$$B(x) = \begin{cases} \infty, & \text{if } g(x) \geq 0 \\ 0, & \text{if } g(x) < 0. \end{cases}$$

Barrier functions

Ideally, $B(x)$ would serve as an indicator function

$$B(x) = \begin{cases} \infty, & \text{if } g(x) \geq 0 \\ 0, & \text{if } g(x) < 0. \end{cases}$$

To avoid numerical issues, the shape of $B(x)$ is controlled by μ.



- As $\mu \to 0$, $B(x)$ becomes closer to an indicator function.
- Notice that $B(x) > 0$ is not required for all feasible points, but must be so for those near $g(x) \leq 0$.

Barrier functions

We will proceed by repeatedly solving $\theta(\mu)$ and iteratively reducing the value of μ. For that to work, we need the following result.

Barrier functions

We will proceed by repeatedly solving $\theta(\mu)$ and iteratively reducing the value of μ. For that to work, we need the following result.

Theorem 1 (Convergence of barrier methods)

Let $f:\mathbb{R}^n \to \mathbb{R}$ and $g:\mathbb{R}^n \to \mathbb{R}$ be continuous functions and $X \in \mathbb{R}^n$ a nonempty closed set in problem P. Suppose $\{x : g(x) < 0, x \in X\}$ is not empty. Let $\overline{x}$ be the optimal solution of P such that, for any neighbourhood $N_\epsilon(\overline{x}) = \{x : ||x - \overline{x}|| \leq \epsilon\}$, there exists $x \in X \cap N_\epsilon$ for which $g(x) < 0$. Then

$$\min\{f(x) : g(x) \leq 0, x \in X\} = \lim_{\mu \to 0^+} \theta(\mu) = \inf_{\mu > 0} \theta(\mu).$$

Letting $\theta(\mu) = f(x_\mu) + \mu B(x_\mu)$, where $B(x)$ is a barrier function observing (1), $x_\mu \in X$, and $g(x_\mu) < 0$, the limit of $\{x_\mu\}$ is optimal to P and $\mu B(x_\mu) \to 0$ as $\mu \to 0^+$.

Outline of this lecture

Barrier functions

Barrier method

Interior point method for LP/ QP

The end... ?

Barrier method

Algorithm Barrier method

```
initialise. ϵ > 0, x^0 ∈ X with g(x^k) < 0, μ^k, β ∈ (0,1), k = 0.
while μ^k B(x^k) > ϵ do
    x̄^{k+1} = arg min { f(x) + μ^k B(x) : x ∈ X }
    μ^{k+1} = β μ^k, k = k + 1
end while
return x^k.
```

Barrier method

Algorithm Barrier method

1: **initialise.** $\epsilon > 0, x^0 \in X$ with $g(x^k) < 0, \mu^k, \beta \in (0,1), k = 0.$
2: **while** $\mu^k B(x^k) > \epsilon$ **do**
3: $\quad \overline{x}^{k+1} = \arg\min \left\{ f(x) + \mu^k B(x) : x \in X \right\}$
4: $\quad \mu^{k+1} = \beta \mu^k$, $k = k + 1$
5: **end while**
6: **return** x^k.

Remarks:

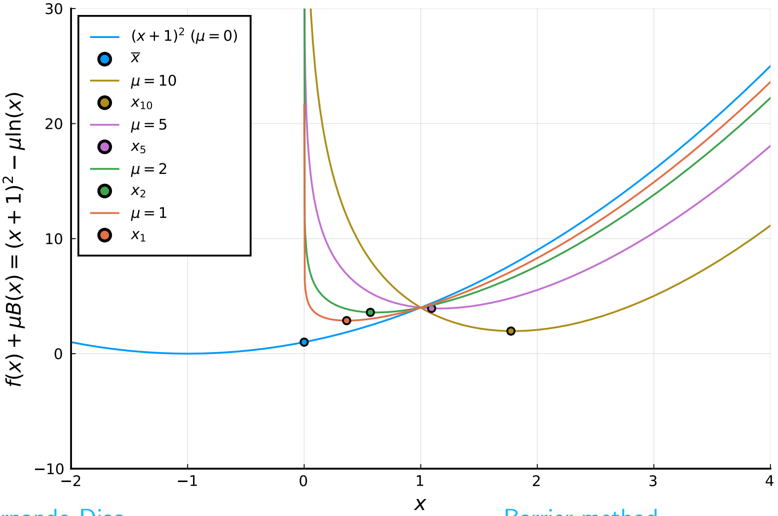
1. Notice that, starting with $x^0 \in X$ with $g(x^0) < 0$, x^k for all $k > 1$ satisfies $g(x^k) < 0$ due to the barrier function.
2. However, applying the barrier method using a fixed step size may cause infeasibility issues.
3. Due to 1., these methods are called interior point methods.

Barrier method

Example 1: $P = \text{min. } \left\{(x+1)^2 : x \geq 0\right\}$ with $B(x) = -\ln(x)$

Barrier method

Example 1: $P = \text{min. } \{(x+1)^2 : x \geq 0\}$ with $B(x) = -\ln(x)$

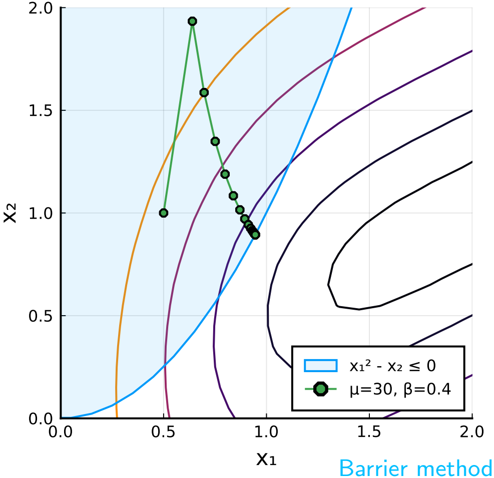


Barrier method

Example 2: $P = \text{min. } \left\{(x_1 - 2)^4 + (x_1 - 2x_2)^2 : x_1^2 - x_2 \leq 0\right\}$ with $B(x) = -\frac{1}{x_1^2 - x_2}$.

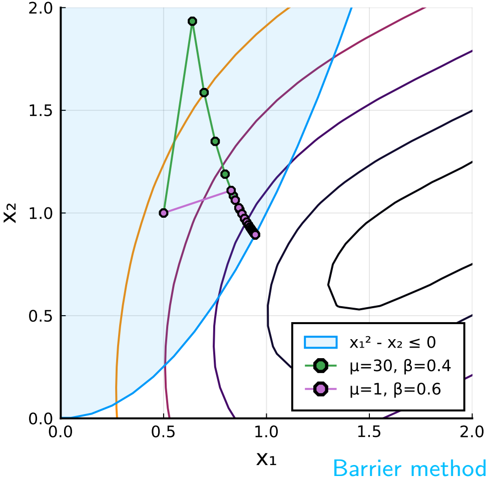
Barrier method

Example 2: $P = \text{min.}\ \left\{(x_1-2)^4 + (x_1-2x_2)^2 : x_1^2 - x_2 \leq 0\right\}$
with $B(x) = -\frac{1}{x_1^2 - x_2}$.



Barrier method

Example 2: $P = \min.\ \left\{(x_1-2)^4 + (x_1-2x_2)^2 : x_1^2 - x_2 \le 0\right\}$
with $B(x) = -\frac{1}{x_1^2 - x_2}$.



Outline of this lecture

Barrier functions

Barrier method

Interior point method for LP/ QP

The end... ?

Interior point methods for LP/QP

Consider the following LP and its dual

$$
\begin{aligned}
(P): \quad \text{min.} \quad & c^\top x \\
\text{subject to:} \quad & Ax = b \; : v \\
& x \geq 0 \; : u
\end{aligned}
\qquad
\begin{aligned}
(D): \quad \text{max.} \quad & b^\top v \\
\text{subject to:} \quad & A^\top v + u = c \\
& u \geq 0, v \in \mathbb{R}^m.
\end{aligned}
$$

Interior point methods for LP/QP

Consider the following LP and its dual

$$(P): \ \text{min.} \ c^\top x \qquad\qquad (D): \ \text{max.} \ b^\top v$$
$$\text{subject to: } Ax = b \ : v \qquad\qquad \text{subject to: } A^\top v + u = c$$
$$x \geq 0 \ : u \qquad\qquad u \geq 0, v \in \mathbb{R}^m.$$

The optimal solution $(\overline{x}, \overline{v}, \overline{u}) = \overline{w}$ satisfies KKT conditions of P:

$$\begin{aligned} &Ax = b, \ x \geq 0 \\ &A^\top v + u = c, \ u \geq 0, \ v \in \mathbb{R}^m \\ &u^\top x = 0. \end{aligned}$$

Let us consider the barrier problem for P by using the logarithmic barrier function.

Interior point methods for LP/QP

The barrier problem is given by:

$$(BP): \text{ min. } c^\top x - \mu \sum_{i=1}^{n} \ln(x_j)$$
$$\text{subject to: } Ax = b.$$

Interior point methods for LP/QP

The barrier problem is given by:

$$(BP): \text{ min. } c^\top x - \mu \sum_{i=1}^{n} \ln(x_j)$$
$$\text{subject to: } Ax = b.$$

The KKT conditions of BP are

$$Ax = b, \ x > 0$$
$$A^\top v = c - \mu \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right).$$

Notice that since $\mu > 0$ and $x > 0$, $u = \mu \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right)$ serve as an estimate for the Lagrangian dual variables.

Interior point methods (IPM) for LP/QP

Let $X \in \mathbb{R}^{n \times n}$ and $U \in \mathbb{R}^{n \times n}$ be defined as

$$X = \mathbf{diag}(x) = \begin{bmatrix} \ddots & & \\ & x_i & \\ & & \ddots \end{bmatrix} \text{ and } U = \mathbf{diag}(u) = \begin{bmatrix} \ddots & & \\ & u_i & \\ & & \ddots \end{bmatrix}$$

and let $e = [1, \ldots, 1]^\top$ be a vector of ones of suitable dimension.

Interior point methods (IPM) for LP/QP

Let $X \in \mathbb{R}^{n\times n}$ and $U \in \mathbb{R}^{n\times n}$ be defined as

$$X = \mathbf{diag}(x) = \begin{bmatrix} \ddots & & \\ & x_i & \\ & & \ddots \end{bmatrix} \text{ and } U = \mathbf{diag}(u) = \begin{bmatrix} \ddots & & \\ & u_i & \\ & & \ddots \end{bmatrix}$$

and let $e = [1, \dots, 1]^\top$ be a vector of ones of suitable dimension.
We can rewrite the KKT conditions of BP as

$$Ax = b,\ x > 0 \tag{2}$$

$$A^\top v + u = c \tag{3}$$

$$u = \mu X^{-1} e \ \Rightarrow\ XUe = \mu e \tag{4}$$

Remark: condition (3) is called relaxed complementarity condition with 0 replaced by μ. This is known as the perturbed KKT system.

Interior point methods (IPM) for LP/QP

According to Theorem 1, $w_\mu = (x_\mu, v_\mu, u_\mu)$ approaches the optimal primal-dual solution of P as $\mu \to 0^+$.

Remarks:

1. The trajectory formed by successive solutions $\{w_\mu\}$ is called a central path due to its interiority forced by the barrier function.

Interior point methods (IPM) for LP/QP

According to Theorem 1, $w_\mu = (x_\mu, v_\mu, u_\mu)$ approaches the optimal primal-dual solution of P as $\mu \to 0^+$.

Remarks:

1. The trajectory formed by successive solutions $\{w_\mu\}$ is called a central path due to its interiority forced by the barrier function.
2. Notice that $XUe = \mu e \Rightarrow u_i x_i = \mu$ for all $i = 1, \ldots, n$.

According to Theorem 1, $w_\mu = (x_\mu, v_\mu, u_\mu)$ approaches the optimal primal-dual solution of P as $\mu \to 0^+$.

Remarks:

1. The trajectory formed by successive solutions $\{w_\mu\}$ is called a central path due to its interiority forced by the barrier function.
2. Notice that $XUe = \mu e \Rightarrow u_i x_i = \mu$ for all $i = 1, \ldots, n$.
3. $c^\top x - b^\top v = u^\top x$ measures the duality gap for the current μ.

Interior point methods (IPM) for LP/QP

According to Theorem 1, $w_\mu = (x_\mu, v_\mu, u_\mu)$ approaches the optimal primal-dual solution of P as $\mu \to 0^+$.

Remarks:

1. The trajectory formed by successive solutions $\{w_\mu\}$ is called a central path due to its interiority forced by the barrier function.
2. Notice that $XUe = \mu e \Rightarrow u_i x_i = \mu$ for all $i = 1, \dots, n$.
3. $c^\top x - b^\top v = u^\top x$ measures the duality gap for the current μ.
4. Also, $u^\top x = \sum_{i=1}^n u_i x_i = n\mu$ is equal to the total slack violation and can be used as a stopping condition.

The notion of interiority

For large enough μ, the solution of the barrier problem is close to the analytic centre of the feasibility set.

The notion of interiority

For large enough μ, the solution of the barrier problem is close to the analytic centre of the feasibility set.

The analytic centre of a polyhedral set $S = \{x \in \mathbb{R}^n : Ax \leq b\}$ is given by the solution of

$$\max_{x} \prod_{i=1}^{m} (b_i - a_i^\top x)$$
$$\text{subject to: } x \in X,$$

i.e., finding $\overline{x} \in S$ of maximum distance to each of the hyperplanes $a_i^\top x = b_i$.

The notion of interiority

For large enough μ, the solution of the barrier problem is close to the analytic centre of the feasibility set.

The analytic centre of a polyhedral set $S = \{x \in \mathbb{R}^n : Ax \leq b\}$ is given by the solution of

$$\begin{aligned} \max_x. \ & \prod_{i=1}^{m} (b_i - a_i^\top x) \\ \text{subject to: } & x \in X, \end{aligned}$$

i.e., finding $\overline{x} \in S$ of maximum distance to each of the hyperplanes $a_i^\top x = b_i$. This is equivalent to

$$\begin{aligned} \min_x. \ & \sum_{i=1}^{m} - \ln(b_i - a_i^\top x) \\ \text{subject to: } & x \in X. \end{aligned}$$

IPM for LP/QP: primal/dual method

Primal/dual path following method is a specialisation of IPM to linear and quadratic problems.

It combines BP with one "additional trick": instead of solving BP to optimality, perform a single Newton step for each μ.

IPM for LP/QP: primal/dual method

Primal/dual path following method is a specialisation of IPM to linear and quadratic problems.

It combines BP with one "additional trick": instead of solving BP to optimality, perform a single Newton step for each μ.

Suppose we start with a $\overline{\mu} > 0$ and a $w^k = (x^k, v^k, u^k)$ sufficiently close to $w_{\overline{\mu}}$. Then, for a sufficiently small $\beta \in (0, 1)$, $\beta\overline{\mu}$ will lead to a w^{k+1} sufficiently close to $w_{\beta\overline{\mu}}$.

Remark: β is typically related to convergence results. Values like $\mu^0 = (x^\top u)/n$ and $\beta \in [0.1, 0.5]$ are often used in practice.

IPM for LP/QP: primal/dual method

Primal/dual path following method is a specialisation of IPM to linear and quadratic problems.

It combines BP with one "additional trick": instead of solving BP to optimality, perform a single Newton step for each μ.

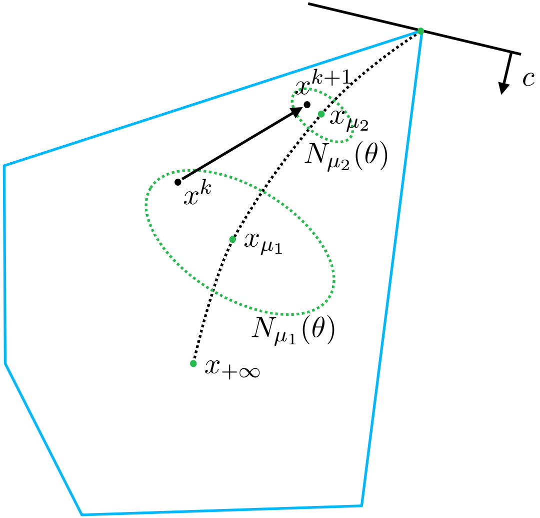
Suppose we start with a $\overline{\mu} > 0$ and a $w^k = (x^k, v^k, u^k)$ sufficiently close to $w_{\overline{\mu}}$. Then, for a sufficiently small $\beta \in (0,1)$, $\beta\overline{\mu}$ will lead to a w^{k+1} sufficiently close to $w_{\beta\overline{\mu}}$.

Remark: β is typically related to convergence results. Values like $\mu^0 = (x^\top u)/n$ and $\beta \in [0.1, 0.5]$ are often used in practice.

For example, let $N_\mu(\theta) = ||X_\mu U_\mu e - \mu e|| \leq \theta\mu$. Then, by selecting $\beta = 1 - \frac{\sigma}{\sqrt{n}}$, $\sigma = \theta = 0.1$, and $\mu^0 = (x^\top u)/n$, successive Newton steps are guaranteed to remain within $N_\mu(\theta)$.

Further reading: see this reference (link) for more details.

IPM for LP/QP: primal/dual method



IPM for LP/QP: primal/dual method

Let the perturbed KKT system (2) – (4) for each $\hat{\mu}$ be denoted as $H(w) = 0$. Let $J(\overline{w})$ be the Jacobian of $H(w)$ at $\overline{w}$.

Applying Newton's method to solve $H(w) = 0$ for $\overline{w}$, we obtain

$$J(\overline{w})d_w = -H(\overline{w}) \tag{5}$$

where $d_w = (w - \overline{w})$.

IPM for LP/QP: primal/dual method

Let the perturbed KKT system (2) – (4) for each $\hat{\mu}$ be denoted as $H(w) = 0$. Let $J(\overline{w})$ be the Jacobian of $H(w)$ at $\overline{w}$.

Applying Newton's method to solve $H(w) = 0$ for $\overline{w}$, we obtain

$$J(\overline{w})d_w = -H(\overline{w}) \tag{5}$$

where $d_w = (w - \overline{w})$. By rewriting $d_w = (d_x, d_v, d_u)$, (5) can be equivalently stated as

$$\begin{aligned} Ad_x &= 0 \\ A^\top d_v + d_u &= 0 \\ \overline{U}d_x + \overline{X}d_u &= \hat{\mu}e - \overline{X}\,\overline{U}e. \end{aligned}$$

The algorithm proceeds by iteratively solving the above system with $\mu^{k+1} = \beta\mu^k$ with $\beta \in (0, 1)$ until $n\mu^k$ is small enough.

IPM for LP/QP: primal/dual method

Remark: Notice that primal feasibility conditions are included in the system $H(w) = 0$. This is typically referred to as the equality constrained Newton's method with "Newton system"

$$\begin{bmatrix} A & 0^\top & 0 \\ 0 & A^\top & I \\ \overline{U} & 0^\top & \overline{X} \end{bmatrix} \begin{bmatrix} d_x \\ d_v \\ d_u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hat{\mu} e - \overline{X}\,\overline{U} e \end{bmatrix}. \tag{6}$$

IPM for LP/QP: primal/dual method

Remark: Notice that primal feasibility conditions are included in the system $H(w) = 0$. This is typically referred to as the equality constrained Newton's method with "Newton system"

$$\begin{bmatrix} A & 0^\top & 0 \\ 0 & A^\top & I \\ \overline{U} & 0^\top & \overline{X} \end{bmatrix} \begin{bmatrix} d_x \\ d_v \\ d_u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hat{\mu} e - \overline{X}\,\overline{U} e \end{bmatrix}. \tag{6}$$

In practice, updates incorporate primal and dual infeasibility, which can be shown to vanish as the algorithm progress. Updates become

$$\begin{bmatrix} A & 0^\top & 0 \\ 0 & A^\top & I \\ U^k & 0^\top & X^k \end{bmatrix} \begin{bmatrix} d_x^{k+1} \\ d_v^{k+1} \\ d_u^{k+1} \end{bmatrix} = - \begin{bmatrix} Ax^k - b \\ A^\top v^k + u^k - c \\ X^k U^k e - \mu^{k+1} e \end{bmatrix}, \tag{7}$$

where $\mu^{k+1} = \beta \mu^k$.

IPM for LP/QP: primal/dual method

Let the residuals (i.e., the amount of infeasibility) be

$$r_p(x,u,v) = Ax - b \text{ (primal)}; \; r_d(x,u,v) = A^\top v + u - c \text{ (dual)}.$$

Let $r(w) = r(x,u,v) = (r_p(x,u,v), r_d(x,u,v))$. The optimality conditions (6) require the residuals to vanish, that is $r(\overline{w}) = 0$.

IPM for LP/QP: primal/dual method

Let the residuals (i.e., the amount of infeasibility) be

$$r_p(x,u,v) = Ax - b \text{ (primal)}; \; r_d(x,u,v) = A^\top v + u - c \text{ (dual)}.$$

Let $r(w) = r(x,u,v) = (r_p(x,u,v), r_d(x,u,v))$. The optimality conditions (6) require the residuals to vanish, that is $r(\overline{w}) = 0$.

Consider the first-order approximation for r at w for a step d_w

$$r(w + d_w) \approx r(w) + Dr(w)d_w,$$

where $Dr(w)$ is the derivative of r evaluated at w. The step d_w for which the residue vanishes is $Dr(w)d_w = -r(w)$ (cf. (7)).

IPM for LP/QP: primal/dual method

Let the residuals (i.e., the amount of infeasibility) be

$$r_p(x,u,v) = Ax - b \text{ (primal)}; \; r_d(x,u,v) = A^\top v + u - c \text{ (dual)}.$$

Let $r(w) = r(x,u,v) = (r_p(x,u,v), r_d(x,u,v))$. The optimality conditions (6) require the residuals to vanish, that is $r(\overline{w}) = 0$.

Consider the first-order approximation for r at w for a step d_w

$$r(w + d_w) \approx r(w) + Dr(w)d_w,$$

where $Dr(w)$ is the derivative of r evaluated at w. The step d_w for which the residue vanishes is $Dr(w)d_w = -r(w)$ (cf. (7)).

The directional derivative of $||r(w + td_w)||_2^2$ in the direction d_w is

$$\frac{d}{dt}||r(w + td_w)||_2^2 \bigg|_{t \to 0^+} = 2r(w)^\top Dr(w)d_w = -2r(w)^\top r(w),$$

which is strictly decreasing.

IPM for LP/QP: primal/dual method

Algorithm Interior point method for LP

1: **initialise.** primal-dual feasible w^k, $\epsilon > 0$, μ^k, $\beta \in (0, 1)$, $k = 0$.
2: **while** $n\mu = c^\top x^k - b^\top v^k > \epsilon$ **do**
3: compute $d_{w^{k+1}} = (d_{x^{k+1}}, d_{v^{k+1}}, d_{u^{k+1}})$ using (7) and w^k.
4: $w^{k+1} = w^k + d_{w^{k+1}}$
5: $\mu^{k+1} = \beta\mu^k$, $k = k + 1$
6: **end while**
7: **return** w^k.

IPM for LP/QP: primal/dual method

Algorithm Interior point method for LP

1: **initialise.** primal-dual feasible w^k, $\epsilon > 0$, μ^k, $\beta \in (0, 1)$, $k = 0$.
2: **while** $n\mu = c^\top x^k - b^\top v^k > \epsilon$ **do**
3:     compute $d_{w^{k+1}} = (d_{x^{k+1}}, d_{v^{k+1}}, d_{u^{k+1}})$ using (7) and w^k.
4:     $w^{k+1} = w^k + d_{w^{k+1}}$
5:     $\mu^{k+1} = \beta\mu^k$, $k = k + 1$
6: **end while**
7: **return** w^k.

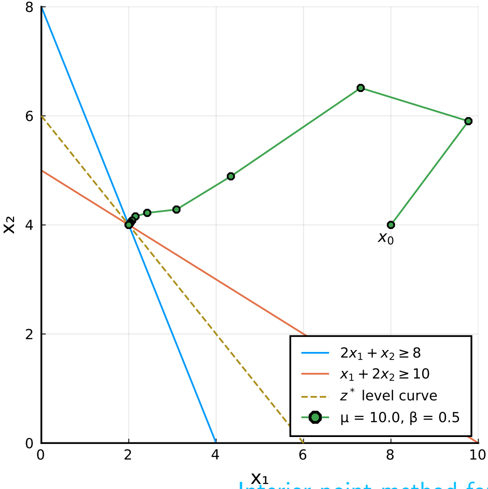
Remarks:

1. Notice that the step size is set to one. A line search could be performed between Lines 3 and 4.
2. This method has polynomial complexity which, under specific conditions, can be shown to be $O(\sqrt{n}\ln(1/\epsilon))$.

The primal-dual interior point (IP) method

Example:

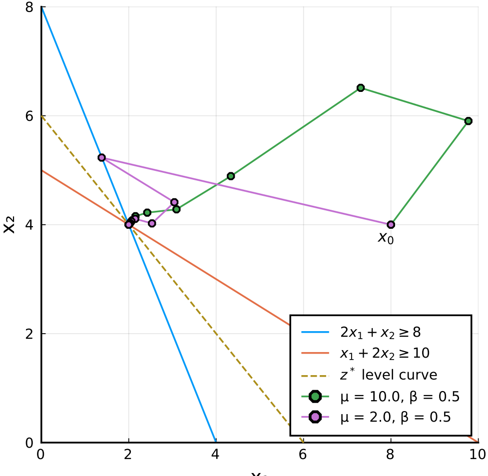
min. $z = x_1 + x_2 : 2x_1 + x_2 \geq 8,\ x_1 + 2x_2 \geq 10,\ x_1, x_2 \geq 0.$



The primal-dual interior point (IP) method

Example:

min. $z = x_1 + x_2 : 2x_1 + x_2 \geq 8,\ x_1 + 2x_2 \geq 10,\ x_1, x_2 \geq 0.$



Outline of this lecture

Barrier functions

Barrier method

Interior point method for LP/ QP

The end... ?

Shameless plug

My research: developing optimisation methods and models (mainly MILP) in different areas such as: air traffic control, bioinformatics and transport.

Find more about my research

Current interests:

- Conflict avoidance and recovery (ACRP);
- Urban air mobility (UAM);
- Flow Decomposition (MFD);
- Transport.

The end... ?


THANK YOU